Routledge
Taylor & Francis Group

ARTICLE

Steering the governance of artificial intelligence: national strategies in perspective

Roxana Radu ⓘ[a,b]

[a]Global Governance Centre, Graduate Institute of International and Development Studies, Geneva, Switzerland; [b]Centre for Socio-Legal Studies, University of Oxford, Oxford, UK

ABSTRACT

As more and more governments release national strategies on artificial intelligence (AI), their priorities and modes of governance become more clear. This study proposes the first comprehensive analysis of national approaches to AI from a hybrid governance perspective, reflecting on the dominant regulatory discourses and the (re)definition of the public-private ordering in the making. It analyses national strategies released between 2017 and 2019, uncovering the plural institutional logics at play and the public-private interaction in the design of AI governance, from the drafting stage to the creation of new oversight institutions. Using qualitative content analysis, the strategies of a dozen countries (as diverse as Canada and China) are explored to determine how a hybrid configuration is set in place. The findings show a predominance of ethics-oriented rather than rule-based systems and a strong preference for functional indetermination as deliberate properties of hybrid AI governance.

Introduction

Artificial intelligence (AI) is the new terrain of contestation in international relations, wrapped in uncertainty about loss of technological control and human oversight. 'Whoever becomes the leader in this sphere will become the ruler of the world', Russian President Vladimir Putin famously stated in 2017 (RT, 2017). Since then, a plethora of public and private actors have issued statements on how AI would change society for the better or for the worse, highlighting infrastructural developments, military applications and impact on jobs and human relations. Some of these statements revealed concrete plans to address AI-related challenges, but the majority remained principled positions on limiting risks associated with disruptive technologies (Ulnicane et al., 2020, Jobin et al., 2019). As recognition grows that tools based on algorithmic processing and machine learning bring about as many promises as commotions, governments are under increased pressure to react for the wellbeing of their citizens and for their *raison d'être* (Taeihagh, 2021). By 2020, more than 30 nations across the globe had started discussions

---

CONTACT Roxana Radu ✉ roxana.radu@graduateinstitute.ch ⬤ Global Governance Centre, Graduate Institute of International and Development Studies, Geneva, Switzerland

about designing national AI strategies. Seventeen of these were already implementing them.

For a long time, the public discourse on AI was closely linked to concepts such as *super-intelligence* (Bostrom, 2014), *technological Singularity* (Shanahan, 2015) and *the Fourth Industrial Revolution* (Schwab, 2017). As the move away from technological perspectives towards societal transformation begins to consolidate, governance approaches start to come under scrutiny (Jobin et al., 2019, Kind, 2020). To date, in both academic and policy writings, the ethics focus has overshadowed the interest in adopting regulation. This article addresses this gap by exploring the following research question: How do states choose to design AI governance arrangements? Through the lens of hybridity, it is argued here that priorities outlined in governmental strategies form the basis for regulatory configurations and functional assignment of roles and responsibilities in policy-making. The drafting process and the national priorities identified thus reflect the extent to which multiple institutional logics inform a hybrid governance approach to AI.

This analysis is based on a qualitative comparison of a dozen national strategies focused on dominant regulatory approaches and the redefinition of public and private roles. The strategies of Canada, China, France, Finland, Germany, Japan, Singapore, South Korea, Sweden, the United Arab Emirates (UAE), the United Kingdom (UK), the United States (US) are examined, providing an overview of the separate discourses of continuity and change in the governance of AI. These national documents embed and reveal mainstream approaches to enabling new markets, regulating emerging technologies and working with non-state actors.

The conceptual framework and the qualitative content analysis presented here make a threefold contribution to governance and public policy literature and debates. First, they bring back the state as a central actor in a field dominated by private governance arrangements, revealing the extent of hybrid interactions in-the-making. Second, they capture the variety of approaches adopted by governments to respond to AI challenges, based on the discourse embedded in sovereigntist AI projects. Thirdly, they uncover the key elements of hybrid governance, pointing out marketization trends and functional indetermination.

The argument unfolds as follows. The next section situates the national strategies against the broader AI debates at the global level, showing that there is a mismatch between the aspiration to place the state in a leading role and the reality of corporations driving developments in the field. This is followed by an in-depth exploration of the hybrid governance literature, outlining merits and limitations of its main tenets. Two dimensions are derived from the hybridity thesis as a way forward in the analysis of AI governance: (1) regulatory approach and (2) redefinition of roles. The third part discusses the findings across the 12 strategies analysed, highlighting broad trends and instances of variation. The conclusion offers a summary of the argument and points to future research directions.

National AI strategies in a private governance landscape

AI was established as a field of research at the 1956 Dartmouth College workshop. It was around that time that undirected research into machine intelligence and robotics

received funding from the American, British and Japanese governments. When public funding was withdrawn in the early 1970s – in what is known as the first 'AI winter' – due to the lack of concrete results and applications, companies stepped in to develop 'expert systems', replicating *if-then* reasoning models in highly specialized areas. The private sector has since taken a leading role in AI research and applications across a wide range of domains, from industrial robots to data mining. In many cases, such advances were first applied in technology industry and later on integrated into general applications, gaining societal acceptance based on earlier successes. As attempts to build a general intelligence supercomputer moved to the background, a successful generation of AI advanced deep learning, speech and image recognition, as well as data analytics tools. They are now deployed on a daily basis in banking, e-learning, medical diagnosis, smart vehicles, etc. forcing governments to formulate a response to the challenges they bring about.

The retreat of the state, first from funding AI research and later on from defining market limitations, has not meant a complete withdrawal of public support; universities and government-sponsored programs continues to conduct relevant research on the topic. But AI garnered extensive support as its application became widespread, culminating in geopolitical tensions and a new 'race to the bottom among powerful nations' (Scharre, 2019). The talk of a 'global AI race' is continuously fuelled by a 'great powers' discourse, with the United States, Russia and China competing for supremacy in the field. The governance of AI is often politicized, revealing concerns that AI could be designed to serve the ideologies and interests of a few centres of power (Jobin et al., 2019, Taeihagh, 2021). This article refocuses the attention on the broader state-society-market interactions in the context of AI, showing the extent to which they form the basis for the transformation of public and private authority.

Around the globe, the growing state interest in AI developments coalesced around the need to control the negative effects and the unintended consequences of new technologies, in particular their impact on furthering inequality. In a largely privatized domain of governance, intergovernmental efforts directed at AI governance have addressed limited aspects of public intervention. In May 2019, OECD member states adopted a set of Principles on AI highlighting democratic values and respect for human rights. Other intergovernmental initiatives stressed the need to work with various stakeholders in order to lead developments in the field. In the Group of Seven, agreement was reached on fostering cooperation with international organisations to promote a human-centred society and to reduce AI-related risks. In March 2018, the Government of Quebec proposed the creation of an *Organisation mondiale de l'intelligence artificielle* as an intergovernmental forum for building consensus on the standards and practices governing the application(s) of AI. While many state-led initiatives have explored the ethical and human rights dimensions of AI (e.g. Council of Europe's Expert Committee on AI and Human Rights, OECD Principles on AI), providing guidelines for future research, they have come short of reforming public governance frameworks and addressing the needs of developing countries.

What has become clear in recent years is that public authorities have supported the dominant focus on AI ethics by putting forward their own recommendations for administrative bodies. The specialized literature also covers proposals around algorithmic impact assessment, laws on AI and robots, as well as fairness, accountability and

transparency frameworks (Cath, 2018). An initial evaluation of 160 sets of AI ethical principles and guideline documents conducted by AlgorithmWatch (2020) showed that the majority of binding agreements and voluntary commitments that exist are proposed by the private sector. A critical take on ethical guidelines designed by companies reveals the oscillation between 'ethics washing', or attempts to disguise ethical stances to avoid regulation, and 'ethics bashing', resulting in loss of hope in the power of normative discussions:

The word 'ethics' is under siege in technology policy. Weaponized in support of deregulation, self-regulation or hands-off governance, 'ethics' is increasingly identified with technology companies' self-regulatory efforts and with shallow appearances of ethical behavior (Bietti, 2019)

While self-regulation has been the main response to AI-enabled transformation at the global level, regulatory discussions and strategic approaches at the regional and national levels have gained prominence since 2016. Stronger critical stances have started to emerge, in particular around the narrow understanding of the responsibilities of AI developers, the monopolization of research by a few companies and the lack of diversity of perspectives in the field (Kind, 2020; AI Now 2018; Cath, 2018). Some scholars have also deplored the limited contextualization of the plethora of AI principles issued by various actors (Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020), while some NGOs pointed out they lacked much-needed enforcement mechanisms (AlgorithmWatch, 2020).

Such concerns resonate with the approach of the European Union (EU) on the matter, materialized in the Ethics Guidelines for Trustworthy AI (April 2019) and the Policy and Investment Recommendations for Trustworthy AI (June 2019). Both documents were issued by the High-Level Expert Group on Artificial Intelligence, a multi-stakeholder group of 52 experts from various sectors. Prominent EU work in this area includes the 2017 European Parliament Resolution on Civil Law Rues on Robotics (non-binding), the 2018 European Commission Communication on Artificial Intelligence for Europe and the 2018 establishment of an advisory European Group on Ethics in Science and New Technologies.

Complementing these efforts are specific approaches to regulating AI-driven innovation – such as autonomous vehicles (Taeihagh & Lim, 2019, Leiman, 2020) – and shaping developments in the field in light of privacy and data protection provisions. In Europe, the General Data Protection Regulation, which imposes significant sanctions for violations, is relevant to AI discussions on two levels: 1) for the collection and storage of data of personal nature and the protection of the rights of data subjects (to access, to object, to rectify, etc.); 2) for enabling the data subject to obtain from the controller information about the logic of the algorithm (Art 15 (1)). The latter is yet to be tested in practice, but creates the basis for a so-called 'right to explanation'.

Through patchwork legislation, the 'return of the state' in the AI field becomes more visible, but only in a fragmented way. All-purpose technology of this magnitude requires broader governance frameworks that restructure basic relations between the public and the private sector, as has been the case with the Internet (Radu, 2019; Radu, Chenou, & Weber, 2014). In this respect, national strategies reveal more clearly the innovations and limits of the public approach taken to govern artificial intelligence.

The countries hosting technology industry giants have taken the lead, with the ambition to dominate AI development at the global level in the next decade. Many other countries, in particular from the developing world, are still debating their national priorities and future AI frameworks. The first country to release a comprehensive national AI strategy in March 2017 was Canada, but the first sector-specific strategy, the one from South Korea, preceded it by one year. At the time of writing (June 2020), AI strategic documents or working group papers were available in more than 17 countries. The mushrooming of strategic initiatives at the national level is likely to continue as more countries discuss their approaches to AI. This article makes a timely contribution to the policy debates in the field, providing a comparison of emerging governance approaches.

So far, states have not imposed strict limitations on AI-related innovation, but that does not mean they have always been passive players. In many cases, they enabled the creation of markets for AI to thrive in. Oftentimes, they funded basic research that led to advancements exploited by businesses. In recent years, governments also started adopting AI technologies for reforming their own administrations. These moves reflect more complex and profound changes in the governance of emerging technologies, that neither the 'retreat of the state' thesis, nor the 'return of the state' thesis capture adequately. The next section presents the hybrid governance tenets and their merits and limitations in relation to AI. It dissects the key elements of hybridity, distilling two dimensions for the empirical analysis.

Theoretical insights: hybridity and AI governance

Designing governance systems for all-purpose technology is not an easy task. Based on the understanding of governance as a process of patterned and orderly interactions between various institutions and actors (Biersteker 2010), institutional arrangements can be examined and responsibilities across the governance spectrum can be disentangled. Where high interaction between public and private actors exists, the hybridity thesis is particularly useful to unpack relationships of mutual dependency in situations of uncertainty (Ménard, 2004), in particular for their functional continuity in achieving public goals (Hodge & Greve, 2005) or performing public responsibilities (Chenou & Radu, 2017; Radu, Zingales, & Calandro, 2015). Sociological institutionalists like Crouch (2005) have long regarded this as the norm rather than the exception in advanced capitalist economies. Denis et al. (2015) note a growing hybridity trend in the public sector, with governments in steering roles or in reactive mode, responding to external pressures. Offering a more nuanced conceptualisation of the interactions and overlaps among governance actors, hybridity sheds light on interests, roles and shared understandings that acquire new institutional forms.

According to Skelcher and Smith (2015, p. 436), hybridity is a 'non-exceptional, but not necessarily universal event'. It can be better explained using the institutional logics approach, which connects normative frames and organizational embodiments in order to identify where agency lies. Hybrid governance not only patches together a multitude of institutional logics, but also blurs the boundaries between institutional forms and actors' identities. Following Friedland and Alford (1991), the organizing principles, material practices and ideological constructions in each society form the basis for institutional logics, understood as 'symbolically grounded, organizationally structured, politically defended, and technically

and materially constrained' (p. 248–49). The influential trilogy of hierarchy, market and networks has been used frequently to differentiate new modes of governance (Thorelli, 1986). In an ideal form, each of these modes denotes an operating logic, corresponding to: public interest, for-profit drive and a mix of goals. Mapping these interests for the field of AI is particularly helpful as a basis for applying hybrid governance tenets.

Drawing on theoretical advances from organizational studies, institutional logics expose interests and priorities specific to each sector, although public-private inter-actions remain widespread and difficult to classify. Cashore et al. (2021) distinguish between these complex and diverse forms of public-private interactions, from coor-dination and collaboration to substitution and co-optation. Structurally-embedded hybridity is thus different from an instance of delegation from public to private actors, or a functional replacement of the latter by the former. The plural logics on which hybrid governance builds ultimately transform the relations of these actors in two respects. First, the negotiation space and their identities become subject to interdependencies (Greenwood, Díaz, Li, & Lorente, 2010; Thornton & Ocasio, 2008). Second, agency itself changes when confronted with a multitude of institu-tional logics, impacting the positioning of key actors and related accountability frameworks. In the words of Skelcher and Smith, 'rather than conceptualizing hybrids descriptively as entities that somehow combine different sectoral character-istics or organizational forms, a theoretically richer approach is to propose that they are carriers of multiple institutional logics' (2015, p. 439).

Adding the institutional logics perspective helps us address the unresolved tensions of hybridity, which continue to be debated, from the fear that everything becomes hybrid (Goodfellow & Lindemann, 2013) to the explanatory power of the concept (Stepputat, 2013). To overcome this limitation, Canclini (1995) suggested starting with the process of hybridization as an ongoing mixing and reconfiguration of sources of authority and power in response to changing political and economic conditions.

If we take the argument further and apply it to AI governance, two dimensions of hybridity need to be examined. The first dimension requires establishing at what stage hybridization emerges and whether it represents a new property of AI governance systems, impacting future policy directions. The comparative analysis below focuses on how hybridity appears, revealing plural institutional logics at work and instances of coherence or incoherence in their underpinning values and ideologies.

A second axis of exploration is the redefinition of actors' roles and identities in hybrid configurations, in response to one of the critiques of hybrid theory, noting the need for better specification. The blurring boundaries between the public and the private may give rise to new institutions and to redefined functions. The ambiguity and uncertainty related to the future of AI may structure interactions as a way to preserve freedom of action. From this perspective, the process of function indetermination reflects a decision to allow for open-ended possibilities: 'tasks so far assigned to the polity can be transposed with increasing ease to a web of "authorities" created for the purpose of making decisions on technical and scientific issues' (Graz, 2006). Whether such changes are entailed in the national strategies is a key part of the analysis.

Methodological considerations

As more and more countries define their approach to AI challenges, strategic documents are issued at various levels, encapsulating the vision of key players. This exploratory analysis of national strategies assesses instances of hybrid governance, with a focus on institutional dynamics and the redefinition of actors' roles and identities. A wide array of working group papers, consultations, guidelines and reports precede and inform the design of a national strategy, but the analysis in this article is limited to governmental strategies or national programmes in their final form – with or without committed resources -, which tend to be carefully worded and thus appropriate for a qualitative content analysis.

Among the national strategies released to date, some are general guiding documents, while others are prescriptive plans with clear priorities and funding attached (Dutton, 2018a). This article analyses the latter, because strategic documents embed collective thinking across different levels of government. A two-step methodology was implemented to select cases and specific dimensions of analysis. Starting from a comprehensive list of all national[1] strategies released by February 2019, those in English and those for which there was official English translation[2] were retained for the analysis. Repositories such as the *Future of Life* and Tim Dutton's (2018b) Overview of National AI Strategies were consulted in order to define the final list of cases. Among these, only the strategies dedicated to AI exclusively, as opposed to AI being listed alongside other digital technologies (e.g. Australia's *Innovation Strategy*, Denmark's *Digital Growth Strategy*), were kept.[3] In the group of countries selected for the analysis, a first distinction could be made between the countries presenting the strategy as guidance for industrial policy (e.g. South Korea, United Kingdom) and those adopting a comprehensive approach merging socio-political and economic incentives (e.g. China, France). After the case selection, 12 national documents were retained for the analysis, detailed in Table 1.

Second, the selected strategies were included in the qualitative content analysis on the dimensions of interest here. Cybersecurity scholars such as Carr (2016) and Weiss and Jankauskas (2018) have drawn attention to how national interest and various understandings of security shape the relationship between government and private actors, providing a useful starting point for grasping a state's general position in contested issue areas. An initial search by keywords pointed to relevant sections in the documents analysed. These included references to 'public', 'authorities', 'state', 'institutions', 'government', 'public interest', 'private sector', 'business', 'industry', 'role', 'responsibility', 'policy'. However, keywords were not enough, as references to dimensions of interest were generally scattered anywhere between the preamble and the action points of the document, thus requiring an in-depth analysis of the context to understand particular formulations and nuances.

Despite the fact that these strategies were released one after the other, their content and approach varied according to the priorities identified, ranging from academic

---

[1]Recognizing that 'national' aggregates different levels, this selection follows the way in which the countries in question refer to their AI strategies, including two federal levels (US, Germany).

[2]Such translations were either provided by the issuing authorities themselves or by NGOs/think tanks and were made available online.

[3]Broader innovation and digital economy strategies generally mention AI and automation in general terms.

Table 1. National strategies included in the analysis.

Date	Name of strategy	Country	Issued by
April 2016	*Artificial Intelligence Information Industry Development Strategy*	South Korea	Korean Ministry of Science, ICT and Future Planning
March 2017	*Pan-Canadian Artificial Intelligence Strategy*	Canada	CIFAR, a Canadian-based charitable organization
May 2017	*AI Technology Strategy*	Japan	Artificial Intelligence Technology Strategy Council, established by the Japanese Prime Minister Shinzō Abe
May 2017	*AI Singapore*	Singapore	Smart Nation and Digital Government Office
May 2017	*Artificial Intelligence Programme*	Finland	Steering group appointed by the Minister of Economic Affairs and Employment
July 2017	*Next Generation AI Development Plan*	China	State Council of China
October 2017	*UAE Strategy for AI*	UAE	UAE Government
March 2018	*AI for Humanity: French Strategy for Artificial Intelligence*	France	French government appointed Cédric Villani to lead the process
March 2018	*AI Sector Deal*	United Kingdom	Department for Business, Energy and Industrial Strategy & Department for Digital, Culture, Media and Sport
May 2018	*National Approach to Artificial Intelligence*	Sweden	Ministry of Enterprise and Innovation
November 2018	*Federal Government's Artificial Intelligence Strategy*	Germany	Federal Ministry for Economic Affairs and Energy, Federal Ministry of Education and Research & Federal Ministry of Labour and Social Affairs
February 2019	*Executive Order on Maintaining American Leadership in Artificial Intelligence/American AI Initiative*	United States	US President Donald Trump

excellence (Canada) and skills development (South Korea) to technological sovereignty (Germany). Finland adopted an explicit nation-wide education focus and prioritized AI services for its public administration. To this aim, the Finnish Ministry of Finance launched #AuroraAI, an autonomous applications network to help create the 'conditions for a people-oriented, proactive society', built around the real-life events of people and business transactions. Collaborations between various AI sub-programmes have also been established, such as in the case of Canada-France-UK partnerships for research, but they are not covered in this study.

Having explained the research design used to assess how configurations of interests and governance decisions come together for defining a national approach to AI, the next section discussed the findings, stressing hybrid governance elements. It delves into broad trends across AI sovereigntist projects and variation in terms of emerging logics, regulatory approaches and newly-created institutions.

Findings

This study set out to explore how states choose to design AI governance arrangements since 2016. Analyses of recent AI strategies released by for-profits and nonprofits globally (Ulnicane et al., 2020) reveal that various forms of tactical engagement begin to replace ad-hoc responses to the disruptive speed and scope of AI transformations. Many national strategies go in the same direction and discuss how to develop and support scientific

research, retain AI talent, and how to enhance skills for future work. Additionally, they all propose – to various extents – the industrialization of AI technologies via sectoral programmes and the uptake of AI by start-ups, and small and medium-sized enterprises. A few countries talk about their ambition to become world leaders in the field (China, the UK, the US) from both a technical and a political standpoint. This section sheds light on broad trends and signification variation in the development and content of national AI strategies, starting with the drafting process.

The national AI strategies analysed here reflect authoritative priorities, directions and allocation of resources that governments have completed in a relative short time span. Yet they vary in scope and length, ranging from sectoral visions of AI development to full-fledged industrial strategies and comprehensive governmental approaches, driven by a multitude of actors. When it comes to the drafting process, a first gap can be noticed between the early adopters of AI strategies and countries in the process of designing their strategies. The first tend to be AI leaders and developed countries, rather than developing countries. Among EU member states, there is coherence around the perceived influence of the block and the need to continue the regional work. While the need for international cooperation is recognized by the majority of countries analysed here, this seems to imply – more often than not – exchanges with countries that are more advanced in AI technologies. The relationship with developing countries is rarely mentioned. One exception is Germany, whose national strategy has an action point around 'building up capacities and knowledge about AI in developing countries in the context of economic cooperation so that economic and social opportunities can be utilized there'.[4]

A second limitation of the drafting process is visible in the diversity of means employed to write a national strategy, in many cases not informed by broad consensus: while some countries hosted long sectoral consultations, others delegated the creation of the strategy to one person (e.g. France) or to a group of experts (e.g. Finland). These processes affected the institutional logics observed and the type of hybridity emerging in the field. Corporate representatives were often driving in the drafting process, as the following engagements show: Jérôme Pesenti (Facebook) contributing to the UK AI Sector deal, the former Nokia CEO Pekka Ala-Pietilä working with the team drafting the Finnish approach to AI, and the former startup entrepreneur and investor Chang Byung-gyu leading the South Korean *Committee of the Fourth Industrial Revolution*, made up of 25 private sector representatives and 5 government officials (Yonhap, 2017).

One important consequence of assigning a dominant place for industry in the drafting process is that hybridity is embedded from the start, without an explicit assumption of power balance between the public and the private sector. Subsequently, in overseeing developments as part of public governance initiatives, industry representatives continue to play an important role, generally constituting at least a third of the members of these bodies. The growing boundary permeability noted by hybrid governance scholars concerns not only sectors, but also practices and, crucially, knowledge. The enduring concentration of emerging technology expertise in private hands (Radu, 2019; Radu et al., 2014) is further accentuated. This complements more advanced commercial strategies of companies driving AI innovation, generally consolidating their position

---

[4]The concluding point in that section of the strategy is that 'developing and emerging economies must not be cut off from technological change'.

around two poles of (digital) power. In international patent applications, China came second after the US in 2018 (WIPO, 2018). A handful of technology companies from these two countries also have the largest AI research investments and presence within the industry bodies developing standards.

National strategies under the magnifying glass: broad trends

The majority of countries included in the analysis embraced a coordinated approach with a centralized vision of AI development. Plural institutional logics are always at play in the strategies analysed here, many times by design. Nowhere is this more visible than in *AI Singapore*, a programme led by the National Research Foundation with participants such as Smart Nation and Digital Government Office, the Economic Development Board, the Infocomm Media Development Authority, the state-owned company SGInnovate, and the Integrated Health Information Systems. Japan provides another example of multiple rationalities coming together, explicitly combining the 'wisdom of industry, academia and government' in order to 'build a framework for sustainable social implementation'. The French strategy talks about strategic engagement in four sectors of particular interest to the state (health, transport/mobility, environment and defence/security), deliberately leaving aside other issue areas such as banking and insurance as 'their development is less a matter of public initiative as it is of private impetus, largely initiated as is, and (...) any State involvement in it would be undesirable'.

Looking at what is missing in these strategies is equally important: from specific details on how the drafting process is conducted to military and surveillance uses of AI. Despite the relevance of emerging technologies for national security, only few countries (China, France and the US) make reference to military interests in the field. Whether explicitly mentioned or not, the pursuit of offensive capabilities underlies a significant part of investments in AI technology.[5] According to the 2019 *AI Global Surveillance Index*, at least 75 countries around the world actively use AI for surveillance, primarily for predictive analysis in smart city platforms, smart policing and facial recognition (Feldstein, 2019).

While many strategies engage in an elaborate, but selective discussion of the changing role of the state, only few problematize the distinction between the public and the private interests (e.g. Finland, Japan, South Korea). The strategic leadership of states in the field remains generally disconnected from the impetus to regulate AI more strictly, although there is consensus around data sharing and standardization across the board, from China to Sweden. The Finnish national programme sees the unlocking of AI potential as dependent on both public and private sectors and declares that 'legislation should naturally also support the change'.

The strong involvement of industry representatives in the expert discussions and AI working groups brings to the surface the perpetuation of functional indetermination. In the absence of an authoritative differentiation, hybrid governance implicitly requires a high reliance on experts via informal mechanisms (Graz, 2014). The mutual influence

---

[5]Internationally, since 2017, the United Nations has been holding expert meetings on AI-directed weapons and has convened a Group of Governmental Experts (as part of the UN Convention on Certain Conventional Weapons) to discuss a ban on lethal autonomous weapon, with 22 of its member states calling for their prohibition. Countries like Israel, Russia, South Korea and the US have opposed this initiative.

of different individuals over each other's decisions through formal and informal rules also characterises notions of pluralistic and non-hierarchical governance present in hybrid configurations.

In the cases analysed here, governments envision broad roles for themselves, such as: leading AI developments worldwide (China, the UK, the US), ensuring technological sovereignty (Germany), overseeing the process of AI adoption (Finland, France), correcting market failures for the most vulnerable (South Korea), or being the first buyer of advanced technologies (UAE). As becomes clear in the Singaporean strategy, the role of the government consists not only in oversight and immediate control, but also in coordinating networks and selecting instruments for policy experimentation. Yet all strategies remain vague on the concrete measures enabling them to act in these roles. Beyond the call for rapidly introducing AI in public administration and modernizing the governmental services in response to the new technological revolution underway, gaps in public investment are often noted with regard to research, human resources, and infrastructure.

This discussion unpacked the logics at play in the development of institutional dynamics and the relatively weak regulatory approach to AI. However, this perspective is incomplete without an exploration of the specific hybrid arrangements set up for governing the field. The interplay between public and private sector is further consolidated in the creation of new institutions, as discussed below.

Variation in the national AI approaches: new roles and institutions

Crucial to the hybridization thesis presented here, the positions and interests of the public and for-profit sectors did not appear to be clearly defined in the national documents analysed, revealing a high degree of functional indetermination. Rather than adopting regulation or presenting a consistent state intervention direction, most governments appear to take a reflexive turn and ponder upon the changes needed. The South Korean strategy deems it 'critical for policymakers to embrace a new technology regulation paradigm and remove regulatory obstacles to innovation' and the German strategy refers to the need 'to factor in the regulatory framework for later use', singling out healthcare as a priority sector.

Instead of a rule-based system, the 12 national strategies introduced and prioritized an ethics orientation. All the documents analysed here (except for the Chinese and American strategies) place an emphasis on designing ethical principles and guiding developments in a normative direction. The German government promoted the use of 'ethics by, in and for design', whereas the French strategy noted that ethical considerations were 'lagging behind practice, but would be necessary for the acceptance of AI'. Whereas the rules and ethics directions are not contradictory and could potentially co-exist, they represent regulatory regimes built on different values and trust systems. The general reluctance to regulate AI at an early stage is reminiscent of the approach chosen for regulating the Internet only when security and legal problems became widespread (Radu, 2019). It is also tightly linked to the fear of stifling innovation and the complex management of uncertainties inherent to new technologies, explicitly mentioned in some of the strategies.

Importantly, the majority of the nations included in this analysis envisioned the creation of special AI Councils or Data Committees to monitor AI adoption and implementation processes. Oversight bodies or AI councils driving the AI policy mandates were generally dominated by representatives of academia and private sector. Overall, NGOs and rights groups were not equally represented (e.g. UK, Canada). AI strategies rarely included end-users of these technologies as a specific group for policy dialogue. The Finnish strategy is an exception here, specifically mentioning that 'cooperation would be needed between the private and public sectors as well as with individual people'. The real-time experimentation specific to AI techniques also applies to the governance dynamics envisioned in the broader politico-economic environment restructured by this technology. This is particularly noticeable in the efforts to create new institutions and new programmes to work on AI.

In designing them, hybridity results in a *common horizon approach* in which market and state actions can no longer be disentangled, as there is a sharing of goals and a growing mutuality and reliance on one another. This may go as far as moving closer to a private sector logic. The French strategy provides an eloquent example: in its proposal to test out sectoral platforms, advantages are weighted and industry moves are mimicked:

'the digital ecosystem is characterized by an omnipresent "winner takes all" logic and dominant positions seem increasingly difficult to challenge. And the fields covered by AI are no exception, which is why it is up to the public authorities to introduce "platformisation" into these various sectors, if only to avoid value being vacuumed off by a private actor in a paramount position'.

This wording shows that different stages can be distinguished in the discourse presented, oscillating between the acknowledgement of the risk of co-optation and the normalization of hybrid elements in the design of future directions for the country.

Diverse approaches regarding the institutionalization of AI governance can be identified in the national strategies. In a few cases (China, Japan, France, the US), existing ministries were asked to drive and coordinate cross-sector work, with responsibilities in the areas they generally cover. France also proposed the creation of a shared specialist centre of 30 members to help provide specific inputs and implement projects in other departments. The UK diverged from this model by proposing a permanent institution, the Office for AI, as a joint unit situated between the Department of Digital, Culture, Media and Sport and the Department for Business, Energy and Industrial Strategy. The UAE went a step further and designated a Minister of State for AI, to oversee technological reforms in the country.

The governments that focused more on research excellence directed their attention to the new institutes and research programmes to be established. Both Canada and Germany envisioned channelling research efforts through existing bodies such as the nonprofit Canadian Institute for Advancement and Research (CIFAR) and the academia and industry-led German Research Centre for Artificial Intelligence. Finland, Sweden and Canada took a broader approach to the development of skills across their societies, enhancing academic leadership and public awareness. France proposed a European DARPA-style organization for AI, alluding to its early success in developing the Internet, but also embracing a Franco-German collaboration via new cross-border cooperation and research centres.

Alongside these structures, many national strategies introduced formal policy input in the form of loosely defined expert groupings (referred to as 'independent' or 'multi-stakeholder'), whose final composition was generally not defined at the time of document publication. They were involved in designing ethical principles for working with AI data and providing broader guidance to the government on AI-related priorities. Examples of these abound: a New Centre for Data Ethics and Innovation and an AI Council in the UK, an AI Consultative Council in the UAE, a Data Ethics Commission in Germany, an Advisory Council on the Ethical Use of AI and Data in Singapore, a Committee for Technological Innovation and Ethics in Sweden. The need for international cooperation in the development of ethical frameworks was explicitly brought forward in the French national strategy, which proposes the establishment of a group of experts beyond national borders.

These newly established bodies represent an institutional response to the uncertainty embedded in new technologies, as their competences are not clearly specified and remain dependent on the members selected. The redefinition of roles for the public and private sectors thus takes a new turn, oscillating between integrating AI governance among the competences of existing ministries, setting up new functions and proposing new bodies with vaguely defined mandates. In most cases, these solutions are combined in what becomes an increasingly complicated configuration for the AI field. Institutionally, there is little variation in the forms of hybridity emerging, confirming the deliberate choice for functional indetermination. Against this background, accountability frameworks remain difficult to establish. What is currently missing in the national strategies is a clear indication of who makes the rules and for how long.

Conclusion

As an all-purpose technology deployed in everyday services, AI requires both national and international governance systems. Current policy debates focus on designing a set of ethical principles in ways that elude the core issues at stake in the new distribution of digital power, beyond a 'race to the bottom' discourse. This study sought to explain how states chose to design AI governance arrangements based on their vision for a national strategy, providing and furthering theoretical insights from hybridization theory. The co-existence of various institutional logics in AI strategies is a property of nascent AI systems which impacts the adoption of a regulatory approach and gives limited attention to developing an accountability framework.

In the 12 national strategies analysed here, uniting the political will and public resources with the industry interests appears to be the preferred recipe for AI policy development. Publicly-funded research with deployments by start-ups and small and medium size companies remains the main strategy of public engagement Although some documents include scarce references to regulation and to the 'red lines' not to be crossed from an ethical perspective, AI industry growth is desired, enabled and facilitated by states all around the world. From China to Germany, very few limitations are imposed on AI development and implementation. Consequently, it becomes increasingly hard to disentangle public interest policies from market dominance interests, a characteristic of plural logic at play in hybrid governance systems.

The emergence of new consultative bodies – with loosely defined mandates – is likely to lead to a greater acceptance of functional indetermination as a mainstream practice for governing AI, which allows for similarly flexible arrangements in the future. It is noteworthy that the interests of governments and industry are closely aligned at the national level. The strong market creation orientation, the vague definition of roles for the public and private sectors, as well as the prioritization of ethical guidelines suggest that hybridity is both an intention of the government and an outcome of the fast AI developments.

Future studies are needed to expand on the effects of the nascent AI ordering, capturing how its early design plays out in the distribution of (digital) power, in particular at the level of private standards and international agendas. Building on this in-depth study of national initiatives, future work should identify the key parameters and dimensions of hybridity permeating the practices of various stakeholders. When it comes to AI innovation, we are reminded that technologically-advanced nations are setting the bar. It is against this background that the AI strategies of developing countries would need to be analysed in the near future.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Notes on contributor

Roxana Radu is a postdoctoral research fellow at the Geneva Transformative Governance Lab and a research associate at the University of Oxford's Centre for Socio-Legal Studies and the Graduate Institute's Global Governance Centre, working on Internet and artificial intelligence governance. She is the author of "Negotiating Internet Governance" (Oxford University Press 2019) and currently serves as the Program Chair of the Global Internet Governance Academic Network (GigaNet).

ORCID

Roxana Radu http://orcid.org/0000-0002-6329-7820

References

AlgorithmWatch. (2020). AI ethics guidelines inventory. Retrieved from https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/

Biersteker, T. (2010). Global governance. In M. Dunn-Cavelty & V. Mauer (Eds.), *The routledge handbook of security studies* (pp. 439–451). London: Routledge.

Bietti, E. (2019). From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy. Final Paper Published in the Proceedings to ACM FAT* Conference (FAT* 2020), Available at SSRN: https://ssrn.com/abstract=3513182

Canclini, N. (1995). *Hybrid cultures: Strategies for entering and leaving modernity*. Minneapolis: University of Minnesota.

Carr, M. (2016). Public–private partnerships in national cyber-security strategies. *International Affairs*, 92(1), 43–62.

Cashore, B., Knudsen, J.S., Moon, J. and van der Ven, H. (2021). Private authority and public policy interactions in global context: Governance spheres for problem solving. *Regulation & Governance*. https://doi.org/10.1111/rego.12395

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions*, 376. doi:10.1098/rsta.2018.0080

Chenou, J. M., & Radu, R. (2017). The 'right to be forgotten': Negotiating public and private ordering in the European Union. *Business & Society*. doi:10.1177/0007650317717720

Crouch, C. (2005). *Capitalist diversity and change: Recombinant governance and institutional entrepreneurs*. Oxford: Oxford University Press.

Denis, J.-L., Ferlie, E., & Van Gestel, N. (2015). Understanding hybridity in public organizations. *Public Administration*, 93(2), 273–289.

Dutton, T. (2018a). Building an AI world: Report on national and regional strategies. *CIFAR*, 6 December. Retrieved from https://www.cifar.ca/cifarnews/2018/12/06/building-an-ai-world-report-on-national-and-regional-ai-strategies

Dutton, T. (2018b). Overview of national AI strategies. *Medium*, 28 June. Retrieved from https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

Feldstein, S. (2019). The global expansion of AI surveillance. 17 September. Carnegie Endowment for International Peace. Retrieved from https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication No. 2020-1, Retrieved from SSRN https://ssrn.com/abstract=3518482

Friedland, R. & Alford, R. (1991). 'Bringing Society Back In: Symbols, Practices and Institutional Contradictions', in W.W. Powell and P.J. DiMaggio (eds), *The New Institutionalism in Organizational Analysis*. Chicago, IL: University of Chicago Press, pp. 232–63.

Goodfellow, T., & Lindemann, S. (2013). The clash of institutions: Traditional authority, conflict and the failure of 'hybridity' in Buganda. *Commonwealth & Comparative Politics*, 51(1), 3.

Graz, J. C. (2006). Hybrids and regulation in the global political economy. *Competition and Change*, 10(2), 230–245.

Graz, J. C. (2014). New players and new processes in global governance: Theorising hybrid governance. Paper presented at the International Political Science Association World Congress in Montreal, Canada, 20–24 July.

Greenwood, R., Díaz, A. M., Li, S. X., & Lorente, J. C. (2010). The multiplicity of institutional logics and the heterogeneity of organizational responses. *Organization Science*, 21(2), 521–639.

Hodge, G. A., & Greve, C. (2005). *The challenge of public–private partnerships: Learning from international experience*. Cheltenham: Edward Elgar.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399.

Kind, C. (2020). The term ethical AI is finally starting to mean something. VentureBeat – The Machine. August 23. Retrieved from https://venturebeat.com/2020/08/23/the-term-ethical-ai-is-finally-starting-to-mean-something/

Leiman, T. (2020). Law and tech collide: Foreseeability, reasonableness and advanced driver assistance systems. Policy & Society, this issue. Retrieved from https://www.tandfonline.com/doi/full/10.1080/14494035.2020.1787696

Ménard, C. (2004). The economics of hybrid organizations. *Journal of Institutional and Theoretical Economics*, 160(3), 345–376.

Radu, R. (2019). *Negotiating internet governance*. Oxford: Oxford University Press.

Radu, R., Chenou, J. M., & Weber, R. (2014). *The evolution of global internet governance: Principles and policies in the making*. Berlin and New York: Springer.

Radu, R., Zingales, N., & Calandro, E. (2015). Crowdsourcing ideas as an emerging form of multistakeholder participation in internet governance. *Policy & Internet*, 7, 362–382.

RT. (2017). 'Whoever leads in AI will rule the world': Putin to Russian children on knowledge day. 1 September. Retrieved from https://www.rt.com/news/401731-ai-rule-world-putin/

Scharre, P. (2019). Killer apps: The real dangers of an AI arms race. *Foreign Affairs*, May/June. Retrieved from https://www.foreignaffairs.com/articles/2019-04-16/killer-apps

Skelcher, C., & Smith, S. R. (2015). Theorizing hybridity: Institutional logics, complex organizations, and actor identities: The case of non-profits. *Public Administration*, 93(3), 433–448.

Stepputat, F. (2013). Contemporary governscapes: Sovereign practice and hybrid orders beyond the center. In M. Bouziane, C. Harders, & A. Hoffman (Eds.), *Local politics and contemporary transformations in the Arab World* (pp. 25–42). UK: Palgrave Macmillan.

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy & Society*. this issue

Taeihagh, A., & Lim, H. S. M. (2019). Governing autonomous vehicles: Emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews*, 39(1), 103–128.

Thorelli, H. B. (1986). Networks: Between markets and hierarchies. *Strategic Management Journal*, 7, 37–51.

Thornton, P. H., & Ocasio, W. (2008). Institutional logics. In R. Greenwood, C. Oliver, K. Sahlin, & R. Suddaby (Eds.), *Handbook of organizational institutionalism* (pp. 99–129). Thousand Oaks, CA: Sage.

Ulnicane, I., Knight, W., Leach, T., Carsten Stahl, B., & Wanjiku, W. G. (2020). Emerging governance for artificial intelligence: Policy frames of government, stakeholders and dialogue. *Policy & Society*. DOI: 10.1080/14494035.2020.1855800.

Weiss, M., & Jankauskas, V. (2018). Securing cyberspace: How states design governance arrangements. *Governance*, 32(2), 259–275.

WIPO. (2018). China drives international patent applications to record heights; Demand rising for trademark and industrial design protection. Press release, 21 March. Retrieved from http://www.wipo.int/pressroom/en/articles/2018/article_0002.html

Yonhap. (2017). Fourth industrial revolution committee to launch by mid-September: ICT minister. *The Korea Herald*. 29 August. Retrieved from http://www.koreaherald.com/view.php?ud=20170829000872